ENHANCED SYSTEM FOR DETECTING INTRUSIONS USING SUPERVISED LEARNING

Suman M[1], Sunil Kumar G[2]

[1] PG Student, Department of Computer Science and Engineering, UVCE, Karnataka, India
[2]Assistant Professor, Department of Computer Science and Engineering, UVCE, Karnataka, India

ABSTRACT

The increasing prevalence of cyber threats in the field of cyber security calls for the creation of strong Intrusion Detection Systems (IDS) in order to protect confidential digital information. With a focus on the NSL KDD Cyber Security dataset, this research offers a novel approach to IDS by using machine learning (ML) and deep learning (DL) approaches. The dataset is well known for providing a thorough representation of network traffic data and is a useful tool for testing and training intrusion detection algorithms. The effectiveness of an extensive number of deep learning (DL), and machine learning (ML), such as Ridge Classifier, K Nearest Neighbour, Nearest Centroid, Decision Tree, Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Multi-layer Perceptron (MLP), Stochastic Gradient Descent (SGD), Regressor Neural Networks, Random Forest, Adaboost, and various neural network architectures, in identifying intrusions within network traffic is assessed in this study. Finding the most effective intrusion detection methods for real-world applications is the aim of this study, advancing cybersecurity procedures and safeguarding vital digital infrastructures via thorough testing and performance assessment.

Keyword : - *Intrusion detection, cyber security, Machine Learning, security, Anomaly detection.*

1. INTRODUCTION-1

The numerous amounts of apps that offer their users services have skyrocketed in the last few decades. Due to the fact that the apps operate on the cloud servers of the service provider rather than the local terminal and both the sources and the outcomes are transmitted via the internet to the users, This kind of service needs little installation work and capacity for processing on the user terminal. Numerous companies have begun to create their streaming services after realizing the clear benefits of providing premium services to customers unable to purchase premium equipment. For example, premium gaming, which usually requires powerful hardware, is now possible to use on any portable device that has reliable internet access. Thanks to exciting services like Google Stadia. Google's processing and rendering of the game real-time user inputs to Google's cloud server, shortly after which the video is transmitted back over the internet to the user's computer. Addition to that, increasing interconnectedness between networks brings about some new attacks to be exploited by criminals. The landscape of cyber security threats a proactive approach safeguarding sensitive information maintaining the integrity of. Furthermore, the advancement of technology further the security landscape making it crucial for to continuously update their defense to stay ahead of threats and breaches, its imperative organizations to be vigilant prepared in addressing these cyber security threats.
The complexity of these threats compounded by the interconnected of modern networks, it’s even more important implement robust security measures Remember, it's just about reacting to cyber, it's about proactive in identifying and mitigating risks to protect your sensitive data and ensure the stability of your network. To defend networks and systems against cyber-attacks and threats, a strong intrusion detection system (IDS) is necessary and it’s vital. The idea of an intrusion detection system was initially carried out by James P. Anderson in 1980s[1].This well-known technology looks for any unauthorized activity or policy infractions on networks. Hence, it is important that IDS technology is constantly improved in order to defend against changing cyber threats and secure sensitive data.

Ultimately, the progress and improvement, Experts in cyber security strongly advised that system administrators should have access to a broad range of tools through the IDS system so they may thoroughly review the network traffic audit trails, the development of an effective and reliable intrusion detection system (IDS) for cyber security has emerged as a pressing problem that has to be resolved in order to thwart these Over the past several years, the complexity of cyberattacks has increased. especially when it comes to cyberattacks on systems that alter or keep confidential data. Claimed that if Understanding the target aberrant activity's network pattern will help. But in order to identify attacks, various strategies have been put forth, including signature-based techniques Histogram-oriented techniques approaches based on Information theory volume. It’s commonly recognized Signature-based techniques are ineffective in handling zero-day attack and new mutants. This is because the signature based method can only identifies the anomalies created in a predefined database of signatures. Clean traffic data is used to create a variety of statistics using histogram-based techniques, and each histogram is placed into a space with the high dimensions. Such methods can easily understandable. Still, false negative rates are frequently quite high. This approach is organized to provide a brief overview of recent studies on machine learning applications in cyber security. The architecture of the created IDS and its parameter options are shown in the model design section. In addition, the specified result section analyses the model with the goal to compare and validate its performance by voting classifier.

This essay’s remaining sections are arranged as follows: In the 2 section, we offers pertinent literature; in the 3 section, we construct the system model and delineate the optimization issue, in the 4 section, the three-phase alternating optimization approach is devised, and in section 5, its effectiveness is evaluated, section 6, finally concludes this endeavor.

2. LITERATURE SURVEY-2

A suggested cybersecurity detection system for intrusion uses a (CNN) convolutional neural network classifier [1], this characteristic sets the suggested IDS model apart from the conventional one. [2] The methodology for anomalous behavior evaluation forms the basis of the system. A profile based on node attributes that is supplied to artificial neural networks that are set up to precisely describe the node's typical actions is part of the recommended technique.[3] hybrid and collective classifiers for machine learning (ML), As machine learning has grown in popularity, researchers and academics have been able to create models of detection systems for intrusion using a variant of classification algorithms.[4] To distinguish between malicious and benign network traffic machine learning system under supervision is created. This model uses a mix of supervised learning techniques and selection of feature approach to determine the best model taking detection success rate into account. The results of the investigation show that, when it came to accurately identifying network traffic with a high detection rate, the design constructed with ANN and cover selection feature performed better. Because of the significant false positive rate of the systems, zero-day attack and novel attack detection is still an active area of study.[5] Applying machine learning to network detection of intrusions has several advantages, one of which is that it requires less expert knowledge than the black or white list model since it uses an extreme learning machine to do balanced constrained optimization. We develop a flexible incremental learning method to find the optimal number of buried neurons. Optimization techniques are devised, along with a customizable way to increase the count of hidden neurons via binary search. for examine the effectiveness for this method, network intrusion detection is used.
[6] Intrusion detection systems come in two flavours: host-based IDS (HIDS) and network-based IDS (NIDS). Host-based detection systems for intrusion are used in networks. admins must remain vigilant about and evaluate activity on a specific computer. In that encrypted data may be viewed while traversing a network, HIDS frequently have this benefit over NIDS. One of its drawbacks is that controlling HIDS is somewhat challenging as each host's configuration and information needs to be managed.[7] It appears that in the near future, machine learning will govern the whole planet. Therefore, we developed the premise that machine learning techniques may be used to help technology-enabled organizations overcome their current issue of recognizing new assaults, or zero day attacks. [8]System For a very long time, anomaly detection has been in great demand. due to instances involving handling unnecessary and inconsequential characteristics in high-dimension data sets.

3. SYSTEM MODEL-3

This part presents the developed system model and the suggested system consists of the learning algorithm and feature selection depicted in Fig. 1. The most pertinent characteristics or attributes must be extracted by the feature selection component in order to identify the example to a particular class or group. The learning algorithm component builds the necessary intelligence or knowledge using the output from the feature selection component. Through training using the training dataset, the model becomes more intelligent. Next, the examining dataset is subjected to the learnt intelligences in order to assess how well the model categorized home much of the unknown material.

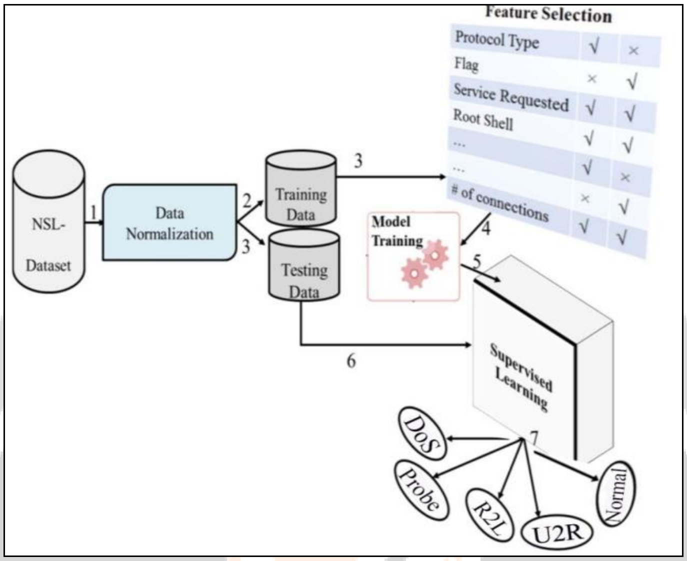


Fig -1: System Model

1. Data Normalization: Normalization is the responsibility of this module. The NSL KDD Cyber Security dataset, which includes tasks such as feature selection, normalization, data cleansing, and categorical variable encoding. The preprocessed dataset will serve as input for the model training phase. 2. Training data: In this module, using the preprocessed data, we will train deep learning and machine learning models. 3. Selection of Feature: The characteristics of the data are chosen in this part. 4. Testing data We'll use a variety of methods, including, decision trees, and neural networks, logistic regression to examine network traffic data and find intrusion. 6. Detection Module: Incoming network traffic will be continually monitored by the real-time detection module. and apply the trained models to identify anomalous or malicious activities. Upon detection of an intrusion, the system will trigger an alerting mechanism to notify system administrators or security personnel. 7. Alerting Mechanism: This module will handle the notification process upon the detection of potential intrusions. System administrators or security personnel will be alerted via email, SMS, or other communication channels. The alert will include relevant information such as the type of intrusion, timestamp, and severity level. Monitoring and Reporting Module: The monitoring and reporting module will provide insights into system performance and detected intrusions. It will continuously monitor system metrics such as detection accuracy, false positives, and response times. Periodic reports summarizing detected intrusions.

4. ALGORITHM -4

1 Data Preprocessing Module:

Task: Prepare the NSL KDD Cyber Security dataset for model training.

Algorithm/Pseudo Code:

```
Step 1: Load the dataset into memory.
Step 2: Perform data cleaning to handle missing values and outliers.
Step 3: Conduct feature selection to identify relevant attributes.
Step 4: Normalize numerical features to a common scale.
Step 5: Encode categorical variables using one-hot encoding or label encoding.
Step 6: break the dataset into two sets for testing and training.
Step 7: Return preprocessed training and testing datasets
```

Model Training Module:

Task: Preprocessed data is used to train Deep Learning (DL) and Machine Learning (ML) models.

Algorithm/Pseudo Code:

```
Step 1: Initialize empty list to store trained models.
Step 2: For each ML algorithm:
 a. Train the algorithm on the preprocessed training data.
 b. Evaluate the model using cross-validation.
 c. Store the trained model in the list.
Step 3: For each DL architecture:
     a. Build and compile the neural network model.
     b. Train the model on the preprocessed training data.
     c. Store the trained model in the list.
Step 4: Return the list of trained ML and DL models.
```

Alerting Mechanism:

Task: Notify system administrators or security personnel upon detection of potential Intrusions.

Algorithm/Pseudo Code:

```
Step 1: Receive notification trigger from real-time detection module.
Step 2: Send alert to designated recipients via email, SMS, or other communication channels.
Step 3: Include relevant information such as intrusion type, timestamp, and severity level in the alert.
Step 4: Log the alert for future reference and analysis.
```

5. EXPERIMENT RESULTS-5

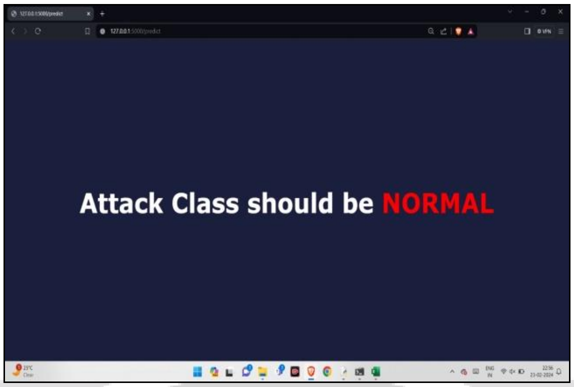

Attack Class should be NORMAL

Fig -2: Predicted Attack Class

On the above screen, the attack is predicted and classified; the output of the prediction is normal, hence no malicious attack.

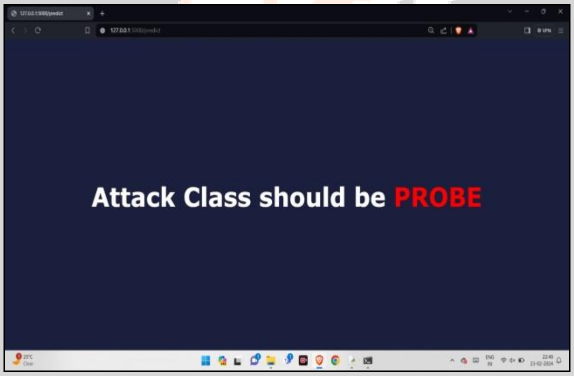

Attack Class should be PROBE

Fig -3: Predicted Attack as Probe

On the above screen, the attack is predicted and classified; the output of the prediction is probe, malicious attack is detected.

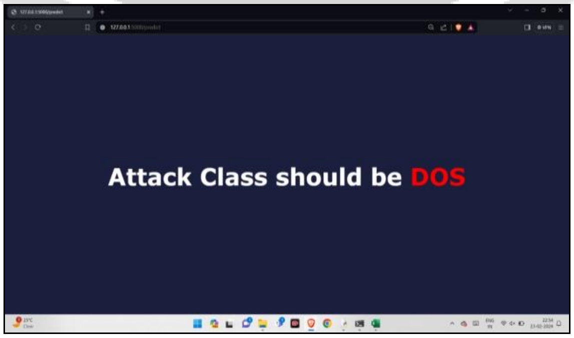
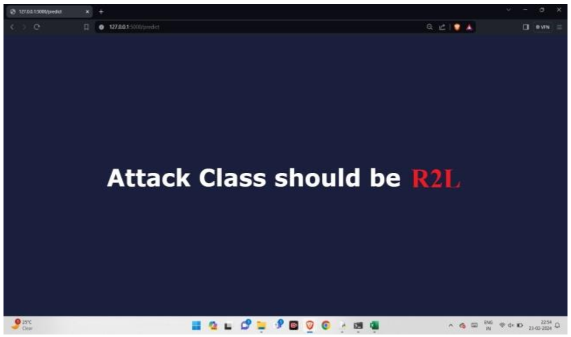

Attack Class should be DOS

Fig -4: Predicted Attack as DoS

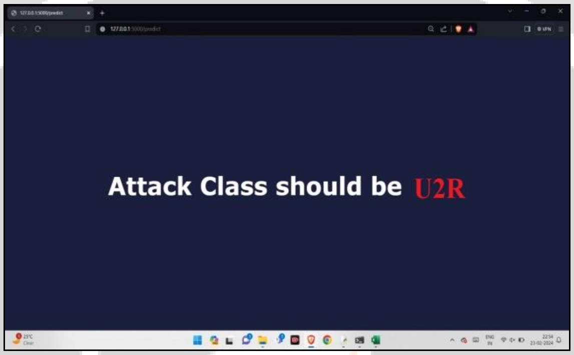
On the above screen, the attack is predicted and classified; the output of the prediction is DoS, malicious attack is detected.



Attack Class should be R2L

Fig -5: Predicted Attack as R2L

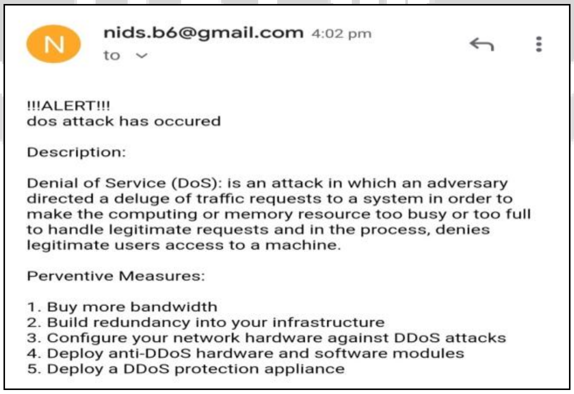
On the above screen, the attack is predicted and classified; the output of the prediction is Root to Local (R2L) attack, malicious attack is detected.



Attack Class should be U2R

Fig -6: Predicted Attack as U2R

On the above screen, the attack is predicted and classified; the output of the prediction is User to Root(U2R) attack, malicious attack is detected.



nids.b6@gmail.com 4:02 pm
to

!!!ALERT!!!
dos attack has occured

Description:

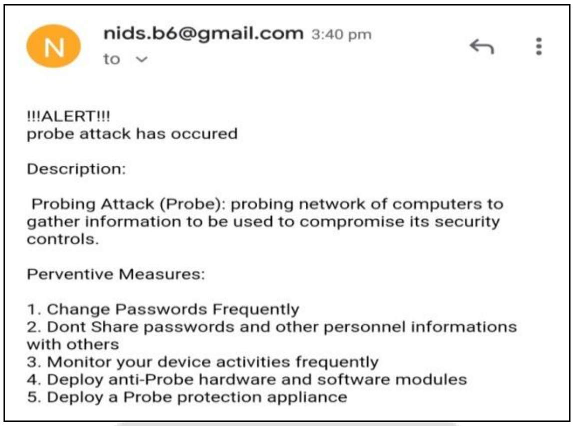
Denial of Service (DoS): is an attack in which an adversary directed a deluge of traffic requests to a system in order to make the computing or memory resource too busy or too full to handle legitimate requests and in the process, denies legitimate users access to a machine.

Perventive Measures:

1. Buy more bandwidth
2. Build redundancy into your infrastructure
3. Configure your network hardware against DDoS attacks
4. Deploy anti-DDoS hardware and software modules
5. Deploy a DDoS protection appliance

Fig -7: Alert Message as Dos

The notification alert message with a description of the attack class and its preventive measures are displayed. The system predicts the attack types as DoS, and the mitigation steps to rectify the attack are given.



nids.b6@gmail.com 3:40 pm
to

!!!ALERT!!!
probe attack has occured

Description:

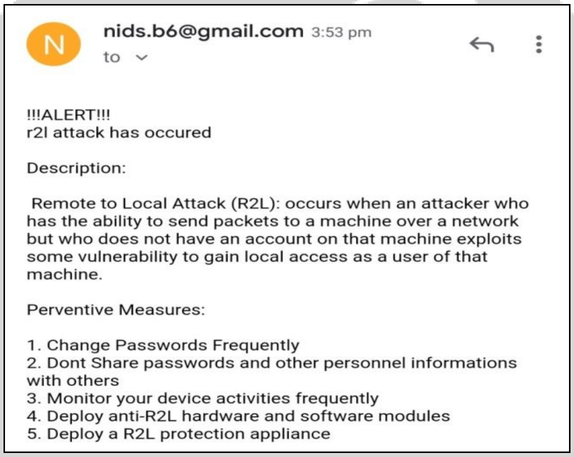
Probing Attack (Probe): probing network of computers to gather information to be used to compromise its security controls.

Perventive Measures:

1. Change Passwords Frequently
2. Dont Share passwords and other personnel informations with others
3. Monitor your device activities frequently
4. Deploy anti-Probe hardware and software modules
5. Deploy a Probe protection appliance

Fig -8: Alert Message as Probe

The system predicts the attack types as Probe, and the mitigation steps to rectify the attack are given.



nids.b6@gmail.com 3:53 pm
to

!!!ALERT!!!
r2l attack has occured

Description:

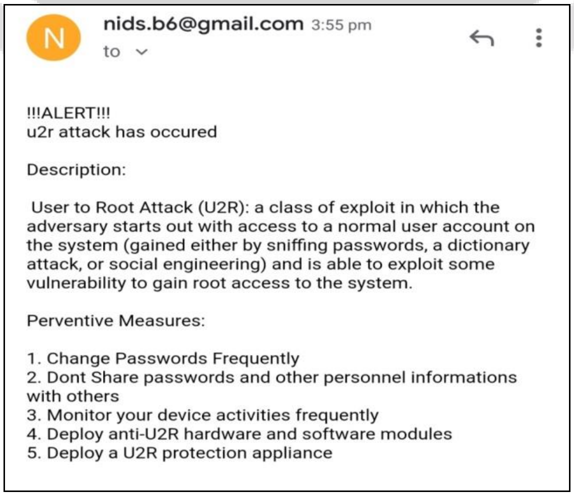
Remote to Local Attack (R2L): occurs when an attacker who has the ability to send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine.

Perventive Measures:

1. Change Passwords Frequently
2. Dont Share passwords and other personnel informations with others
3. Monitor your device activities frequently
4. Deploy anti-R2L hardware and software modules
5. Deploy a R2L protection appliance

Fig -9: Alert Message as R2L

The system predicts the attack types as U2R, and the mitigation steps to rectify the attack are given.



nids.b6@gmail.com 3:55 pm
to

!!!ALERT!!!
u2r attack has occured

Description:

User to Root Attack (U2R): a class of exploit in which the adversary starts out with access to a normal user account on the system (gained either by sniffing passwords, a dictionary attack, or social engineering) and is able to exploit some vulnerability to gain root access to the system.

Perventive Measures:

1. Change Passwords Frequently
2. Dont Share passwords and other personnel informations with others
3. Monitor your device activities frequently
4. Deploy anti-U2R hardware and software modules
5. Deploy a U2R protection appliance

Fig -10: Alert Message as U2R

The system predicts the attack types as U2R, and the mitigation steps to rectify the attack are given.

6. CONCLUSIONS

In conclusion, the advancement of the Enhanced Intrusion Detection System Using Supervised Learning project represents a significant step towards bolstering cyber security measures and safeguarding critical digital assets against the ever-evolving landscape of security threats. By using cutting-edge machine learning techniques (ML) and Deep Learning (DL) techniques, coupled with meticulous data preprocessing and real-time detection capabilities, the system demonstrates its efficacy in detecting both known and novel intrusion attempts with high accuracy. Furthermore, the integration of alerting mechanisms and monitoring/reporting functionalities ensures timely responses to security threats and facilitates ongoing evaluation and refinement of the system's performance. As cyber threats continue to proliferate, the IDS project stands as a robust and adaptive solution, capable of providing organizations with the necessary tools and insights to mitigate risks and protect against potential breaches.

7. REFERENCES

[1] A. JESUS PACHECO, “Artificial Neural Networks-Based Intrusion Detection System for Internet of Things Fog Nodes”, vol. 8, pp. 73907–73911, 2020.

[2] SAMSON HO , SALEH AL JUFOUT “A Novel Intrusion Detection Model for Detecting Known and Innovative Cyberattacks Using Convolutional Neural Network.” Vol. 2, pp. 14-19, 2020.

[3] Usman Shuaibu Musa Megha Chhabra, “Intrusion Detection System using Machine Learning Techniques: A Review.” vol.90, pp. 149–154, 2020.

[4] Kazi Abu Taher, Billal Mohammed Yasin Jisan, Md. Mahbubur Rahman,“ Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection,” vol.978, pp. 643-646, 2019.

[5] Chie-Hong Lee, Yann-Yean Su, Yu-Chun Lin and Shie-Jue Lee, “Machine Learning Based Network Intrusion Detection,” vol.978,pp. 78-83, 2017.

[6]Shalini G, Jaya Kumar M, Abhishek P, Dhamodaran M, “A Network Intrusion Detection System Using Supervised Learning Techniques,”vol.1,pp.1-5, 2020.

[7] Anish Halimaa A, Dr. K.Sundarakantham, “Machine Learning Based Intrusion Detection System,” vol.32, pp. 916–920, 2019.

[8] Arshid Ali, Shahtaj Shaukat, Muhammad Tayyab , Muazzam A Khan, Jan Sher Khan “Network Intrusion Detection Leveraging Machine Learning and Feature Selection” vol.17,pp.49-53, 2020.

[9] L. Bottou, F. E. Curtis, and J. Nocedal, “Optimization methods for large-scale machine learning,” vol. 60, no. 2, pp. 223–311, 2018.

[10] S.Yeom and K. Kim, “Detail analysis on machine learning based malicious network traffic classification.” in Proc. Int. Conf. Smart Media Appl., pp. 49–53. pp. 1–3, 2019.

[11] Y. Gu, K. Li, Z. Guo, and Y. Wang, “Semi-supervised K-means DDoS detection method using hybrid feature selection algorithm,” IEEE Access, vol. 7, pp. 64351–64365, 2019.

[12] O. Faker and E. Dogdu, “Intrusion detection using big data and deep learning techniques,” in Proc. ACM Southeast Conf., , pp. 86–93. 2019.

[13] A. Ahmim et al., “A novel hierarchical intrusion detection system based on decision tree and rules-based models,” in Proc. IEEE 15th Int. Conf 2019.

[14] M. R. Mohammadi, S. A. Sadrossadat, M. G. Mortazavi, and B. Nouri, ‘‘A brief review over neural network modeling techniques,’’ in Proc. IEEE Int. Conf. Power, Control, Signals Instrum. Eng. (ICPCSI), pp. 54–57. 2017.

[15] P. Satam, H. Alipour, Y. Al-Nashif, and S. Hariri, ‘‘Anomaly behavior analysis of DNS protocol,’’ J. Internet Serv. Inf. Secur., vol. 5, no. 4, pp. 85–97, 2015.

[16] S. Fayssal, S. Hariri, and Y. Al-Nashif, ‘‘Anomaly-based behavior analysis of wireless network security,’’ in Proc. 4th Annu. Int. Conf. Mobile Ubiquitous Syst., Netw. Services (MobiQuitous), , pp. 1–8. 2007.

[17] W. –C. Lin, Shih-Wen K. Chih-Fong “Intrusion detection system based on combining cluster centers and nearest neighbors. Knowledge-Based Systems” 78 (pp. 13-21). Elsevier. (2015).

[18] Bolon –C.V. (2012) Feature Selection and Classification in Multiple class datasets-An application to KDD Cup 99 dataset. https://doi.org/10.1016/j.eswa.2010.

[17] Huang, G.-B., Zhou, H., Ding, X., & Zhang, R."Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics" 42(2), 513–529, 2012.
[18] H.Wang,J.Gu,andS.Wang,''An effective intrusion detection framework based on SVM with feature augmentation,'' Knowl.-Based Syst.,vol.136,pp.130–139, Nov. 2017.
[18] Wathiq Laftah Al-Yaseen , Zulaiha Ali Othman , Mohd Zakree Ahmad Nazri; "Multi-Level Hybrid Support Vector Machine and Extreme Learning Machine Based on Modified K-means for Intrusion Detection System", ELSEVIER, Expert System with Applications, Volume.66, ,pp.296-303,Jan 2017.
[19] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "A system for denial-of-service attack detection based on multivariate correlation analysis," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 2, pp. 447–456, 2014.
W. Meng, W. Li, and L.-F. Kwok, "Efm: [20]Enhancing the performance of signature-based network intrusion detection systems using enhanced filter mechanism, Computers & Security, vol. 43, pp. 189–204, 2014.
[21] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in Web Research (ICWR), 2017 3th International Conference on, , pp. 178–184. 2017.
[22] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in International Conference on Networked Systems,, pp. 513–517. 2015.
[23] C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89, pp. 117–123, 2016.
[24] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy, 2018, pp. 108–116.
[25] R. Panigrahi and S. Borah, "A detailed analysis of CICIDS2017 dataset for designing intrusion detection systems," Int. J. Eng. Technol., vol. 7. no. 24, pp. 479–482, 2018.
[26] V. Sze et al., "Efficient processing of deep neural networks: A tutorial and survey," Proc. IEEE, vol. 105, no. 12, pp. 2295–2329, 2017.
[27] A. Paszke et al., "Automatic differentiation in pytorch," in Proc. 31st Conf. Neural Inf. Process. Syst., Long Beach, CA, USA, pp.1-4, 2017.
[28] J. Zheng, F. Shen, H. Fan, and J. Zhao, "An online incremental learning support vector machine for large-scale data," Neural Computing and Applications, vol. 22, no. 5, pp. 1023–1035, 2013.
[29] F. Gharibian and A. A. Ghorbani, "Comparative study of supervised machine learning techniques for intrusion detection," in Communication Networks and Services Research, 2007. CNSR'07. Fifth Annual Conference on, pp. 350–358, 2007.
[30] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85–117, 2015.
[31] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in Military Communications and Information Systems Conference (MilCIS), pp. 1–6. 2015.
[32] T. Janarthanan and S. Zargari, "Feature selection in UNSW-NB15 and KDDCUP'99 datasets," in Industrial Electronics (ISIE), 2017 IEEE 26th International Symposium on, pp. 1881–1886,2017.
[33] L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 6, pp. 446–452, 2015.